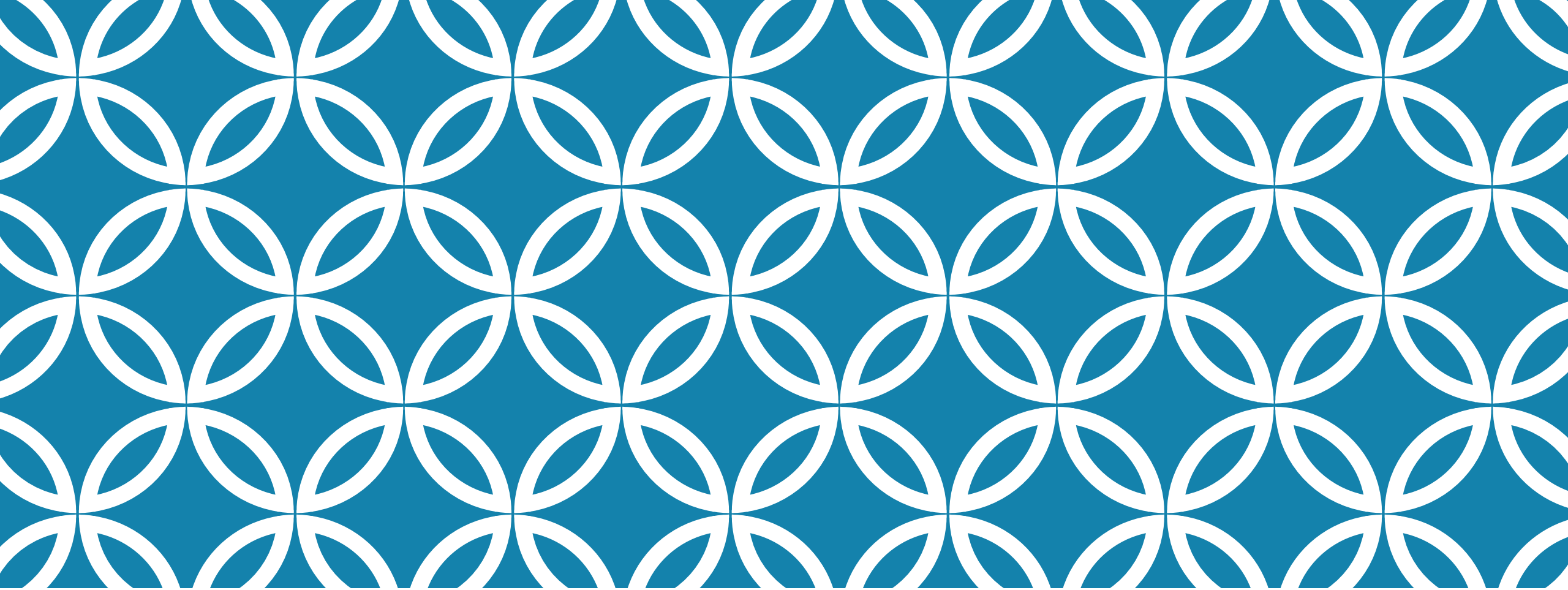


CAPSTONE BATTLE OF NEIGHBOURHOODS

Clustering Neighbourhoods of Mumbai & Delhi

AGENDA

- Background and business Problem
- Data Acquisition
- Methodology
- Result
- Discussion & Conclusion

BACKGROUND & BUSINESS PROBLEM

Mumbai and Delhi are the two most important metro cities in India. There has always been a comparison in terms of quality of life, jobs, education, entertainment and recreational facilities that these cities have to offer to its residents.

This project elaborates on a data science project that attempts to analyse the neighbourhoods in each of these two cities and tries to understand what is popular in them and what they have to offer to someone who is contemplating to make a choice about living in either of the metro cities.

Business problem:

People who would be interested in this study are those who would like to know about potential life and activities in these metro city neighbourhoods if they move to live in one of them. The decision to choose one over the other would depend on popular venues in the neighbourhoods in each of these metro cities.

DATA ACQUISITION

- For this study, we will need data about neighbourhoods in each of these metro cities. The data published by the government on postal codes for all India would serve us well for this study. We will specifically download the CSV provided under https://data.gov.in/resources/all-india-pincode-directory-contact-details-along-latitude-and-longitude.
- In this study, we will download the CSV, read it into a pandas Data frame and curate it to remove the data related to all other cities, towns, and places which are not Mumbai or Delhi, since we are only interested in comparing these two biggest metro cities in India.
- We shall then clean up the unnecessary columns in the CSV, which is not relevant or useful for our current study. Post office names (office name) will be used as the neighbourhood names in each of the regions such as Mumbai or Delhi.
- Neighbourhood names with the same Pin code will be combined as a single row.
- Foursquare API will be used to find the longitude and latitude of each of the neighbourhoods in both Mumbai and Delhi. This will form the dataset we will use for this study.
- Combining these sources of data we will cluster the neighbourhoods to group them and understand them in depth.

METHODOLOGY

Acquiring Postal Codes data:

- First step in the project is to get a list of all neighbourhoods or areas in the country. This data we obtained from the data posted by the Indian govt. on a official website. The file was downloaded and read into a data frame.
- The data was filtered to take only the data pertaining to Delhi & Mumbai. Also, the columns that were not necessary for our analysis was removed.
- Neighbourhoods were grouped into a single row separated by comma delimiter to get the data frame as shown.

	pincode	regionname	Neighborhood
0	110001	Delhi	Baroda House S.O, Bengali Market S.O, Bhagat S...
1	110002	Delhi	A.G.C.R. S.O, Ajmeri Gate Extn. S.O, Darya Gan...
2	110003	Delhi	Delhi High Court Extension Counter S.O, Delhi ...
3	110004	Delhi	Rashtrapati Bhawan S.O
4	110005	Delhi	Anand Parbat Indl. Area S.O, Anand Parbat S.O,...

METHODOLOGY

Acquiring Latitude & Longitude Details

- Pgeocode library was installed. It was used to collect latitude & longitude details for the postal codes in the country 'IN' using the function 'query_postal_code' by giving the postal codes from our earlier as input. The columns not necessary for the analysis was removed to give the following data frame.
- The 2 data frames were merged on Pin code column.
- The data was divided into 2 data frames, one for Delhi and one for Mumbai. Data frame for Delhi is shown below.

	pincode	regionname	Neighborhood	latitude	longitude
0	110001	Delhi	Baroda House S.O, Bengali Market S.O, Bhagat S...	28.6369	77.218229
1	110002	Delhi	A.G.C.R. S.O, Ajmeri Gate Extn. S.O, Darya Gan...	28.6453	77.245600
2	110003	Delhi	Delhi High Court Extension Counter S.O, Delhi ...	28.5947	77.225270
3	110004	Delhi	Rashtrapati Bhawan S.O	28.6453	77.212800
4	110005	Delhi	Anand Parbat Indl. Area S.O, Anand Parbat S.O,...	28.6551	77.188775

METHODOLOGY

Acquiring Venue details

- Using the Foursquare API, we collect the details of venues in each of the neighbourhoods by creating a function getNearbyVenue().
- A new data frame was created by grouping the data based on each neighbourhood for both Mumbai and Delhi data.
- A function was created to get the 10 most common venues for Mumbai and Delhi neighbourhoods separately.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	A I Staff Colony S.O, Santacruz P&t Colony S.O	Café	Coffee Shop	Modern European Restaurant	Asian Restaurant	Indian Restaurant	Tea Room	Spa	Yoga Studio	Farmers Market	Farm
1	Aareymilk Colony S.O, Nagari Niwara S.O, S R P...	Smoke Shop	Bookstore	Indian Restaurant	Bakery	Fast Food Restaurant	Department Store	Dessert Shop	Dhaba	Diner	Dumpling Restaurant
2	Agashi S.O, Chikhal Dongre B.O, Kophrad B.O, V...	Dhaba	Yoga Studio	Convenience Store	Deli / Bodega	Department Store	Dessert Shop	Diner	Dumpling Restaurant	Electronics Store	Farm
3	Agripada S.O, BPC Jacob Circle S.O, Chinchpok...	History Museum	Multiplex	Fast Food Restaurant	Snack Place	Fish & Chips Shop	Cosmetics Shop	Deli / Bodega	Department Store	Dessert Shop	Dhaba
4	Airoli B.O, Airoli S.O	Pizza Place	Toy / Game Store	Café	Dumpling Restaurant	Asian Restaurant	Fast Food Restaurant	Gym	Hotel Bar	Dessert Shop	Fish & Chips Shop

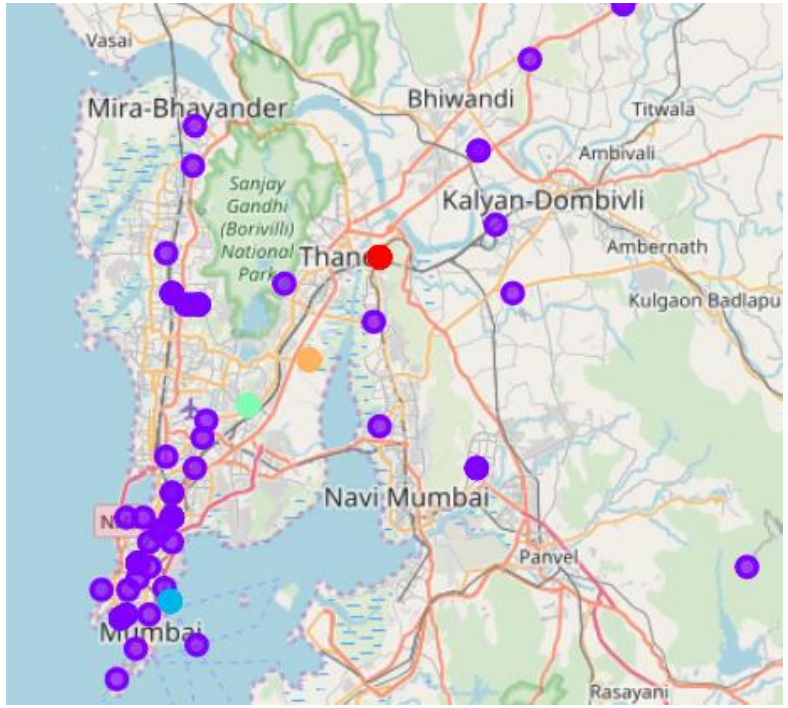
CLUSTERING NEIGHBOURHOODS

Using K Means algorithm, the neighbourhoods in Mumbai and Delhi were clustered into 5 clusters. We try to find the similarities and differences based on the results.

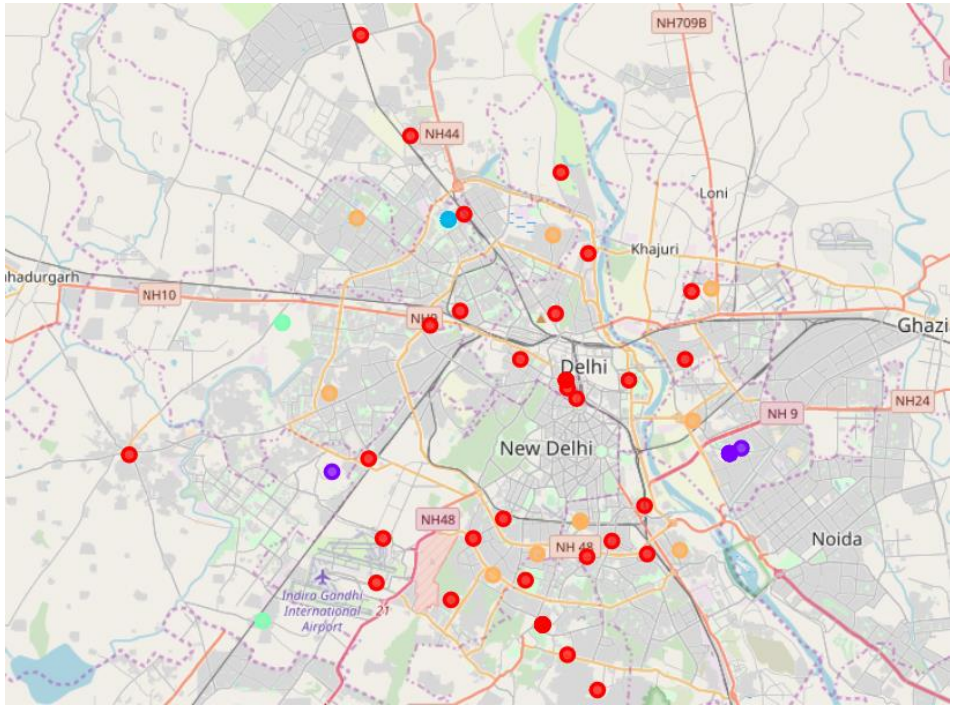
We will further discuss the details in the next sections.

RESULT

Mumbai Clusters



Delhi Clusters



MUMBAI CLUSTERS

Cluster 0:

12 neighbourhoods belong to this cluster. The neighbourhoods belonging to this cluster are popular for having Fish and Chips shop, Yoga Studio, Delis, and Convenience stores. This neighbourhood seems like a place where a young population would enjoy as it comes with various restaurants of different cuisines as well Yoga Studios and necessities like convenience store.

Cluster 1:

76 neighbourhoods are grouped into this cluster. This is the largest cluster in Mumbai. This Cluster contains harbour, gym, ATM, cafes, flea markets as well as variety of stores and restaurants.

Cluster 2:

5 neighbourhoods are grouped into this cluster. The most common venue in this cluster is boat or ferry. So, it would be a good place for people who depend on boats or ferries for transport.

Cluster 3:

6 neighbourhoods are grouped into this cluster. The most common venue in this cluster is Men's store followed by different restaurants.

Cluster 4 :

7 neighbourhoods are grouped into this cluster. This Cluster contains Movie theatres as the 2nd most common venues. Could be a good place for theatre buffs to hang around.

DELHI CLUSTERS

Cluster 0:

36 neighbourhoods belong to this cluster. This cluster has pubs, coffee shops, restaurants, in addition to malls and History Museums.

Cluster 1:

8 neighbourhoods belong to this cluster. This cluster mainly contains IT services, Gym and snack places. It might be an area to stay for employees working in IT services.

Cluster 2:

3 neighbourhoods belong to this cluster. This cluster is a good shopping hub as it contains shoe stores, women's and mobile phone stores, fabric shop along with dessert and donut shops.

Cluster 3:

3 neighbourhoods belong to this cluster. This cluster mainly contains ATM, gyms and gardens. Might be a good place for joggers.

Cluster 4 :

12 neighbourhoods belong to this cluster. The neighbourhoods belonging to this cluster is popular for having Indian restaurants. It also contains Athletics & sports venues and modern European restaurants.

DISCUSSION & CONCLUSION

In this project, we have taken the data for two of India's metro cities and have tried to analyse the neighbourhood regions in these metro cities based on the top venues they have. We have clustered the neighbourhoods based on the most common top venues in each of the neighbourhood. We tried to understand the difference in the type of venues in these metros, which can offer decision points for anybody who is considering settling in either of the metro cities.

Given our cluster information for both Mumbai and Delhi, we see that **Mumbai** and its neighbourhoods are a great place for a **foodie**. There are a lot of restaurants, cafes, bars, etc in Mumbai neighbourhoods. Also due to the proximity of Mumbai to the seashore, Mumbai neighbourhoods offer for **harbours**, **seafood**, boat, and ferry rides.

On the other hand, **Delhi** neighbourhoods and good for those who like **Arts and Crafts**, **History Museums** and Pizza places. There is very less in terms of foreign cuisine restaurants in Delhi.

Mumbai, on the other hand, is great for international visitors, expats, etc, because of the variety and types of food outlets it has. Delhi is inland and its neighbourhoods have proximity to Water Parks, Museums and Arts, and Crafts stores.